IMDB Movie Analysis

PROJECT DESCRIPTION:

The IMDB Movie Analysis project aims to explore and analyse a comprehensive dataset of movies available on the IMDB platform. This dataset contains essential information about movies, including director names, movie titles, duration, genre, budget, gross earnings, IMDb ratings, and more. Through in-depth data analysis using Excel, Data Visualization and Statistics techniques this project seeks to extract valuable insights and trends that contribute to a movie's success.

In this project, I was required to provide a detailed report for the below data record mentioning the answers of the questions that follows:

A. **Movie Genre Analysis:** Analyse the distribution of movie genres and their effect on the IMDB score.

- **Task:** Determine the most common genres of movies in the dataset. Then, for each genre, calculate descriptive statistics (mean, median, mode, range, variance, standard deviation) of the IMDB scores.

B. **Movie Duration Analysis:** Analyse the distribution of movie durations and its effect on the IMDB score.

- **Task:** Analyse the distribution of movie durations and identify the relationship between movie duration and IMDB score.

C. **Language Analysis:** Situation: Examine the distribution of movies based on their language.

- **Task:** Determine the most common languages used in movies and analyse their effect on the IMDB score using descriptive statistics.

D. **Director Analysis:** Effect of directors on movie ratings.

- **Task:** Identify the top directors based on their average IMDB score and analyse their contribution to the success of movies using percentile calculations.

E. **Budget Analysis:** Explore the relationship between movie budgets and their financial success.

- **Task:** Analyse the correlation between movie budgets and gross earnings, and identify the movies with the highest profit margin

-

My Approach:

I have gone through the dataset and understood all the given columns. Then I have observed that there are total of 28 columns and 5031 rows. This dataset also contains unwanted columns. Null Values and blank rows. So, I decided to clean the dataset for analysis:

1) I deleted the columns which is not required for our analysis and left out with only 9 columns which are Director's Name, Duration , Movie Title , genre, Budget , gross , imdb_rating, Language and country.
2) Then selected all the blank cells using go to special, select all and delete entire rows with blank cells.
3) Search for duplicate data and deleted all the duplicates,
4) Removed the Â from all the movies names using find and replace.

Now we're left with 3854 rows and 9 Columns.
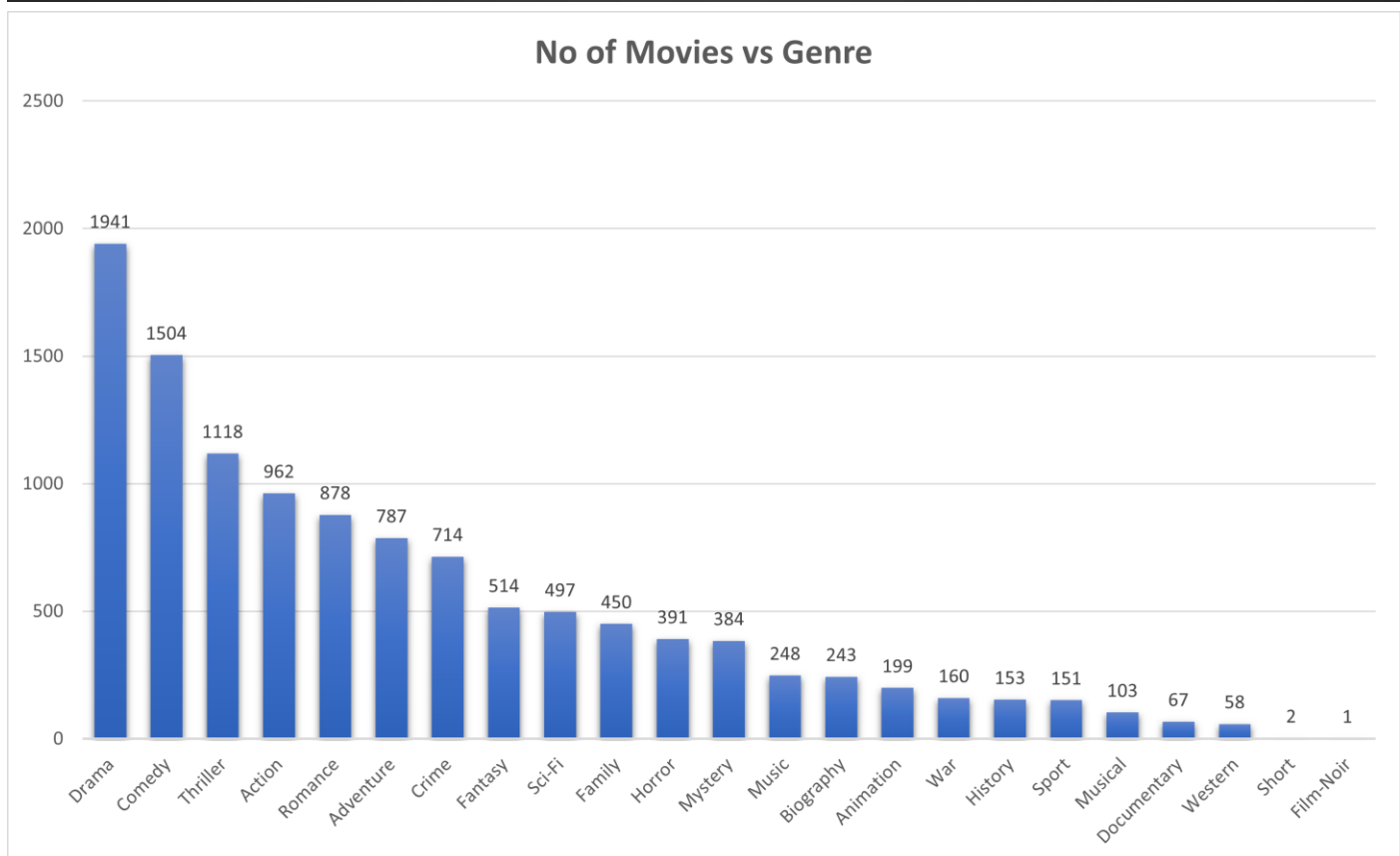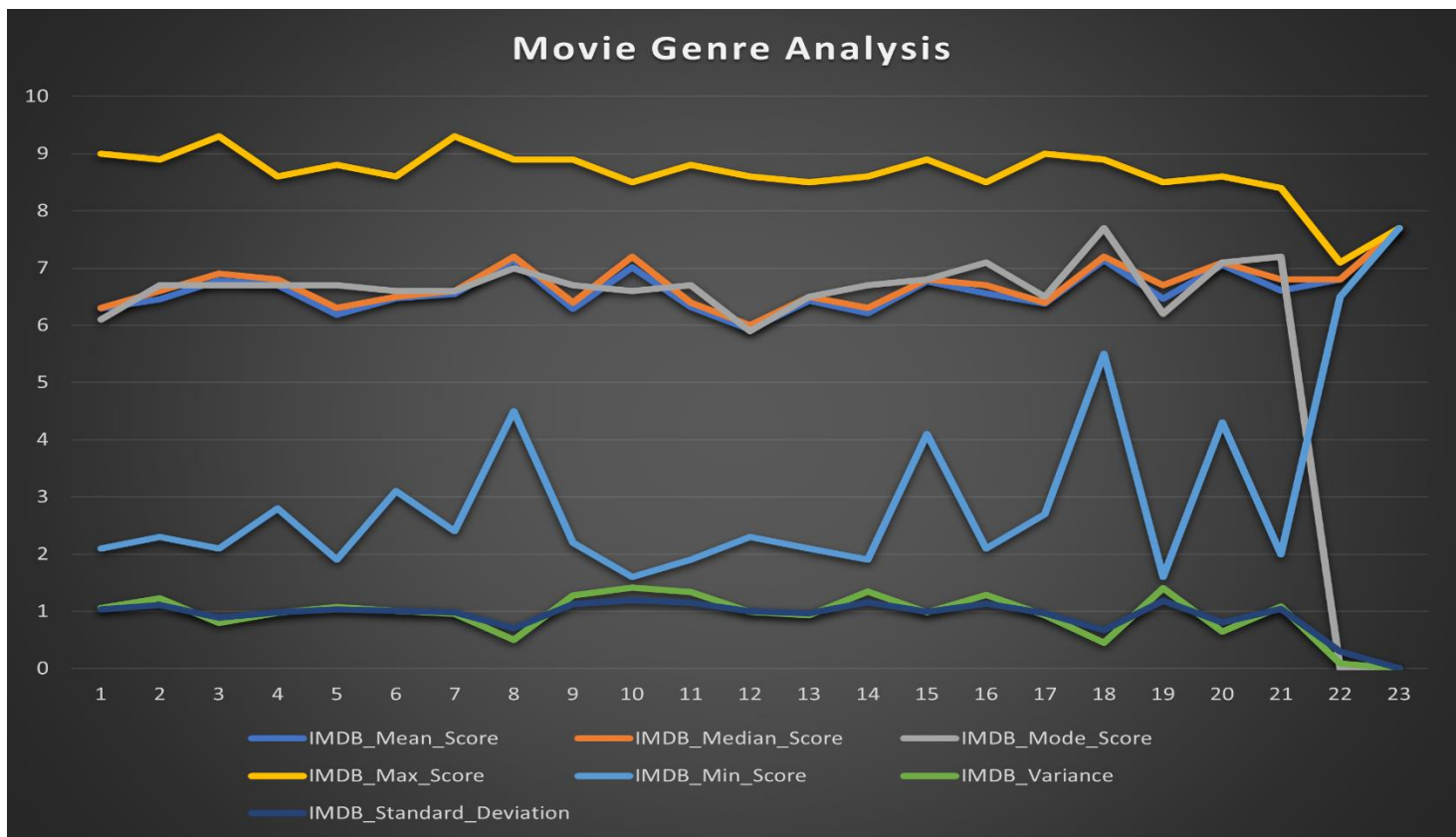
Tech Stack:

For this project I used Microsoft excel 365 to run functions and get answers for the above questions and also graphs.

Insights :

A. **Movie Genre Analysis:** Analyse the distribution of movie genres and their effect on the IMDB score.

- **Task:** Determine the most common genres of movies in the dataset. Then, for each genre, calculate descriptive statistics (mean, median, mode, range, variance, standard deviation) of the IMDB scores.

Genres	No of Movies	IMDB_Mean_Score	IMDB_Median_Score	IMDB_Mode_Score	IMDB_Max_Score	IMDB_Min_Score	IMDB_Variance	IMDB_Standard_Deviation
Action	962	6.290748441	6.3	6.1	9	2.1	1.062648296	1.030848338
Adventure	787	6.458322745	6.6	6.7	8.9	2.3	1.226420567	1.107438742
Drama	1941	6.786192684	6.9	6.7	9.3	2.1	0.793647586	0.890869006
Animation	199	6.700502513	6.8	6.7	8.6	2.8	0.972411808	0.98610943
Comedy	1504	6.184441489	6.3	6.7	8.8	1.9	1.075981337	1.037295202
Mystery	384	6.472916667	6.5	6.6	8.6	3.1	1.008641493	1.004311452
Crime	714	6.544817927	6.6	6.6	9.3	2.4	0.959363902	0.979471236
Biography	243	7.141563786	7.2	7	8.9	4.5	0.500618131	0.707543731
Fantasy	514	6.290077821	6.4	6.7	8.9	2.2	1.27412334	1.128770721
Documentary	67	7.011940299	7.2	6.6	8.5	1.6	1.418364892	1.190951255
Sci-Fi	497	6.322736419	6.4	6.7	8.8	1.9	1.333869373	1.154932627
Horror	391	5.926086957	6	5.9	8.6	2.3	0.991646836	0.99581466
Romance	878	6.42881549	6.5	6.5	8.5	2.1	0.933600191	0.966229885
Family	450	6.210444444	6.3	6.7	8.6	1.9	1.348579802	1.161283687
Western	58	6.765517241	6.8	6.8	8.9	4.1	0.979845422	0.989871417
Musical	103	6.559223301	6.7	7.1	8.5	2.1	1.289211047	1.135434299
Thriller	1118	6.377370304	6.4	6.5	9	2.7	0.932341206	0.965578172
History	153	7.136601307	7.2	7.7	8.9	5.5	0.449510018	0.670455082
Music	248	6.460483871	6.7	6.2	8.5	1.6	1.40553525	1.185552719
War	160	7.05375	7.1	7.1	8.6	4.3	0.648735937	0.805441455
Sport	151	6.603311258	6.8	7.2	8.4	2	1.081379764	1.039894112
Short	2	6.8	6.8	0	7.1	6.5	0.09	0.3
Film-Noir	1	7.7	7.7	0	7.7	7.7	0	0

Movie Genre Analysis

Legend: IMDB_Mean_Score, IMDB_Median_Score, IMDB_Mode_Score, IMDB_Max_Score, IMDB_Min_Score, IMDB_Variance, IMDB_Standard_Deviation

No of Movies vs Genre

| Genre	No of Movies
Drama	1941
Comedy	1504
Thriller	1118
Action	962
Romance	878
Adventure	787
Crime	714
Fantasy	514
Sci-Fi	497
Family	450
Horror	391
Mystery	384
Music	248
Biography	243
Animation	199
War	160
History	153
Sport	151
Musical	103
Documentary	67
Western	58
Short	2
Film-Noir	1

The Most common movie genres from the dataset are Drama, Comedy, Thriller and Action.

B. **Movie Duration Analysis:** Analyse the distribution of movie durations and its effect on the IMDB score.

- Task: Analyse the distribution of movie durations and identify the relationship between movie duration and IMDB score.

Operations	Values
Mean	109.6647059
Median	105
Mode	101
Variance	516.8857876
Standard Deviation	22.73512233

Duration impact on IMDB Score
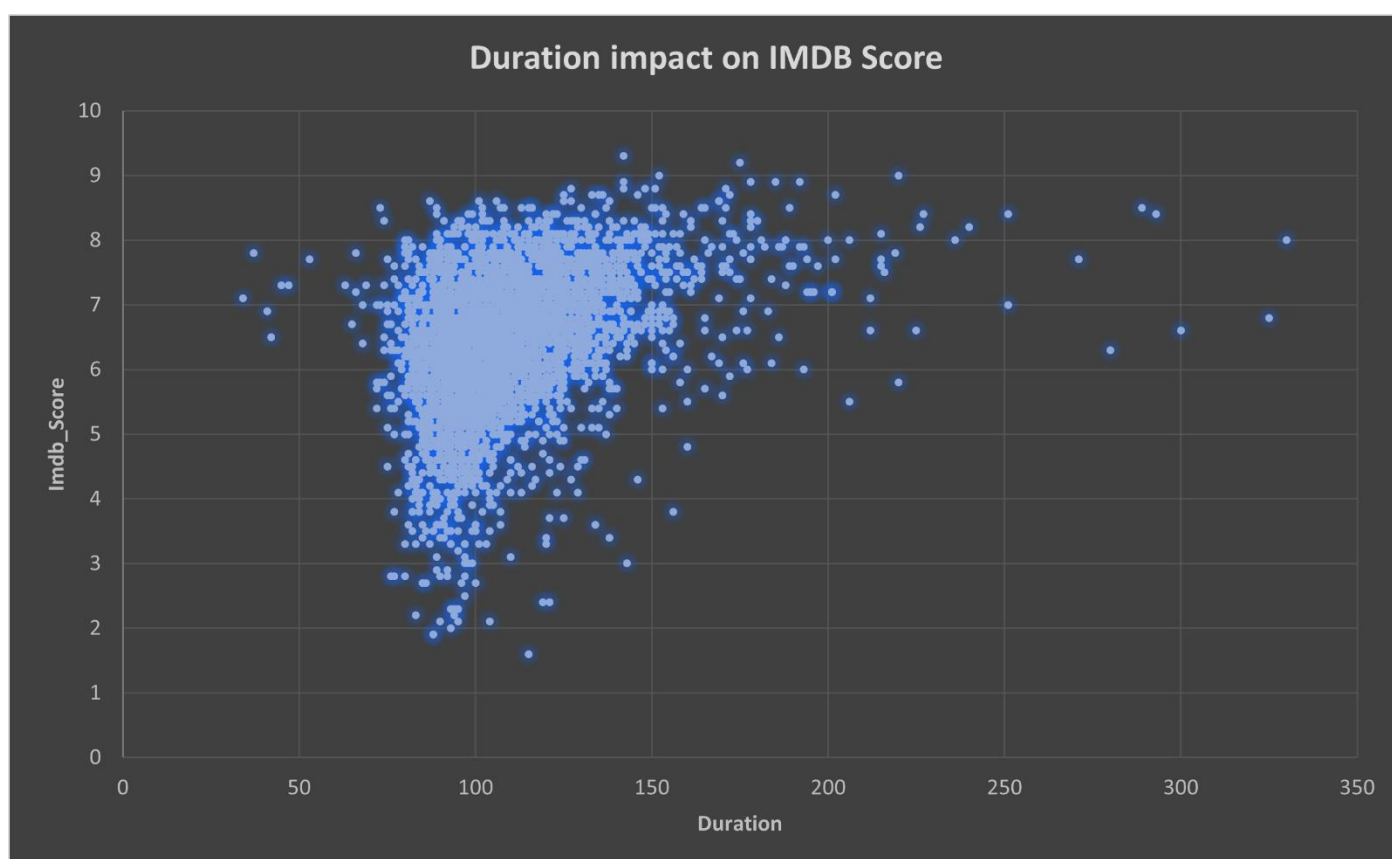
The Average duration of a Movie is 109 minutes.

The movies have high average IMDB rating which are between 100-150 mins.

C. **Language Analysis:** Situation: Examine the distribution of movies based on their language.

- **Task:** Determine the most common languages used in movies and analyse their effect on the IMDB score using descriptive statistics.

Language	No Of Movies	imdb mean	imdb_median	imdb_mode	imdb_variance	imdb_standard_deviation
English	3729	6.41	6.5	6.7	1.123405775	1.05990838
French	37	7.29	7.2	7.2	0.306574142	0.553691378
Spanish	27	7.04	7.1	7.2	0.632839506	0.795512103
Dari	16	7.5	7.4	7.6	0.502773438	0.709065186
Mandarin	14	7.02	7.25	7.6	0.544540816	0.737930089
German	13	7.69	7.7	7.4	0.379171598	0.615769111
Japanese	12	7.63	7.8	0	0.741875	0.861321659
Hindi	10	6.76	7.05	0	1.1124	1.05470375
Cantonese	8	7.24	7.3	7.3	0.16984375	0.412121038
Italian	7	7.19	7	0	1.144081633	1.069617517

The Most common languages used in the movies are English, French, Spanish, Dari, Mandarin and German. I have also Observed that the languages Telugu and Persian have the highest average IMDB score

D. **Director Analysis:** Effect of directors on movie ratings.

- Task: Identify the top directors based on their average IMDB score and analyse their contribution to the success of movies using percentile calculations.

Director Name	No OF Movies	Average Imdb Score	Percentile
Tony Kaye	1	8.6	0.998
Charles Chaplin	1	8.6	0.998
Alfred Hitchcock	1	8.5	0.996
Ron Fricke	1	8.5	0.996
Damien Chazelle	1	8.5	0.996
Majid Majidi	1	8.5	0.996
Christopher Nolan	8	8.43	0.995
Sergio Leone	3	8.43	0.995
S.S. Rajamouli	1	8.4	0.993
Richard Marquand	1	8.4	0.993
Asghar Farhadi	1	8.4	0.993
Marius A. Markevicius	1	8.4	0.993
Lee Unkrich	1	8.3	0.991
Fritz Lang	1	8.3	0.991
Lenny Abrahamson	1	8.3	0.991
Billy Wilder	1	8.3	0.991
Pete Docter	3	8.23	0.99
Hayao Miyazaki	4	8.23	0.99

I have identified that Tony Kaye, Charles Chaplin, Alfred Hitchcock, Ron Fricke, Damien Chazelle, Majid Majidi, Sergio Leone, Christopher Nolan, SS Rajamouli, Richard Marquand, Asghar Farhadi and Marius A Markevicius are the top 10 directors with average IMDB score >=8.4.

E. **Budget Analysis:** Explore the relationship between movie budgets and their financial success.
- **Task:** Analyse the correlation between movie budgets and gross earnings, and identify the movies with the highest profit margin.

Movie Title	budget	Profit in Millions
Avatar	237000000	$52,35,05,847
Jurassic World	150000000	$50,21,77,271
Titanic	200000000	$45,86,72,302
Star Wars: Episode IV - A New Hope	11000000	$44,99,35,665
E.T. the Extra-Terrestrial	10500000	$42,44,49,459
The Avengers	220000000	$40,32,79,547
The Lion King	45000000	$37,77,83,777
Star Wars: Episode I - The Phantom Menace	115000000	$35,95,44,677
The Dark Knight	185000000	$34,83,16,061
The Hunger Games	78000000	$32,99,99,255
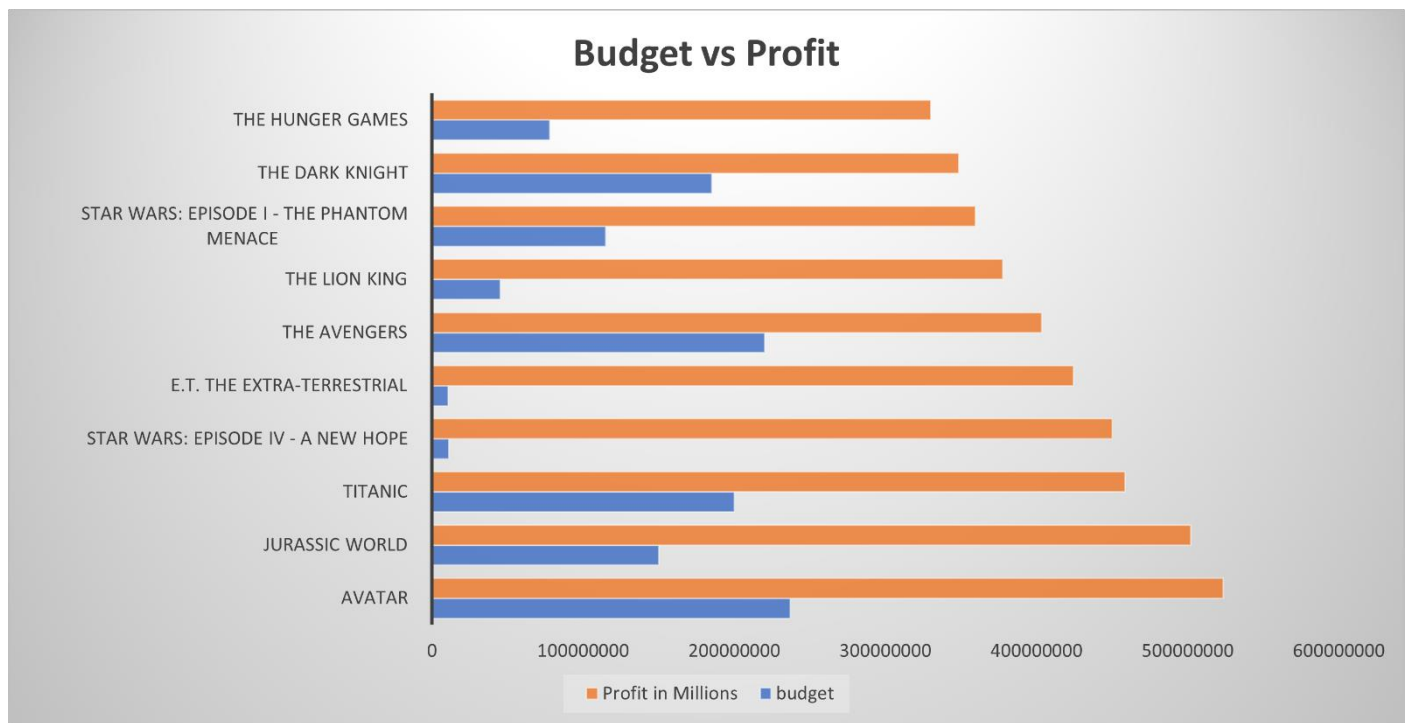
The Top-5 with highest profits are Avatar, Jurassic World, Titanic, Star Wars: Episode IV - A New Hope and E.T. The Extra-Terrestrial. The Correlation between budget and gross is positive.

RESULTS :

With the help of this project, I have gained valuable skill for data analysis using statistical knowledge and excel data visualization. Through this, I have learnt to apply my data analysis skills in solving real life problems